

Collaborative Model Development for Vocabulary and Guidelines

Diane E. Oliver, MD,^{1,2} Edward H. Shortliffe, MD, PhD,² and the InterMed Collaboratory^{2,3,4,5}
¹Palo Alto VA, ²Stanford University, ³Columbia University,
⁴Harvard University, and ⁵University of Utah

The InterMed Collaboratory is a group of medical informatics laboratories, all of which have an interest in controlled vocabularies and in delivery of patient-care guidelines. Each group has worked on these topics for years and therefore has a local perspective on the issues involved in design and development of controlled vocabularies and clinical guidelines. Each group has also learned that creation of site-specific vocabularies, guidelines, and the software systems that make them accessible is time-consuming and expert-intensive. Although the requirements placed on vocabularies and guidelines vary depending on an institution's needs, core generic models for the representation of controlled vocabularies and patient-care guidelines could potentially be shared. If such models were robust and sensitive to cross-application requirements, they could form the basic structures upon which site-specific vocabularies and guidelines could be built, and thus shared among developers.

The term "model" in this context refers to the critical aspects of representation and of desired input-output functionalities to which implementations are expected to conform. Specific implementations of such models on central networked servers then allow vocabularies and guidelines to be entered, stored, accessed, modified, and shared among sites. If two different implementations conform to the same underlying model, it will make it easier to translate between them, to build programming interfaces between them, and to facilitate distributed development of multi-authored content.

In order to identify features that are common to multiple sites and implementations, an approach is to study existing systems, to analyze their similarities, and to strive to reach consensus. This is the approach being undertaken by the InterMed Collaboratory.

With the goal of building common models, and understanding the needs and perspectives of colleagues at other sites, the InterMed participants have communicated using the following methods: 1) sending email, 2) posting documents on the World Wide Web, 3) sharing programs and files via the Internet, 4) scheduling biweekly telephone conferences, 5) arranging face-to-face meetings among members at individual sites, and 6) scheduling occasional face-to-face meetings among InterMed participants from different sites.

The process of model development begins with the exchange of ideas to distinguish those details of structure and function that are general and sharable from those that are local or implementation-specific. Participants present their ideas and receive feedback from the others. Over time, an understanding of the views and requirements of different sites is developed. That which is general and sharable can form the basis for a common model, and an appropriate representation for the model is created. Functionality must be clarified so that new prototypes or links between existing systems can then be built.

The InterMed Collaboratory has made preliminary progress in developing a common model for controlled vocabularies and a common model for guidelines. The model for controlled vocabularies has been influenced by systems at Columbia (the Medical Entities Dictionary [MED]), Stanford (T-HELPER), Brigham & Women's Hospital (BWH), Massachusetts General Hospital (MGH; COSTAR), and the University of Utah (LOINC). In addition, all participants have been influenced by well-known vocabularies such as MeSH, UMLS, ICD-9-CM, and SNOMED. For representation of vocabularies, Columbia uses its internal representation of the MED, Stanford uses Ontolingua, MGH uses MUMPS, BWH uses Thenetsys, and the University of Utah uses a relational database.

The model for guidelines has been based on the experiences of Columbia, Stanford, BWH, and MGH. These institutions have done prior work with guidelines such as tuberculosis screening and management, preventive medicine, breast mass work-up, and oncology and AIDS clinical trial protocols. Representation systems used include the Arden Syntax at Columbia, GEODE-CM at BWH, MBTA at MGH, and PROTÉGÉ at Stanford.

Given the similar yet diverse backgrounds of multiple sites with common goals, frequent communication among these sites, and the belief that sharing expertise is mutually beneficial, it is possible to develop common models. This is being shown with the development of common models for controlled vocabularies and patient-care guidelines. Ultimately software utilizing these common models may be shared as well.